Fine-Grained Censorship Mapping
Information Sources, Legality and Ethics

Joss Wright
Oxford Internet Institute
joss.wright@oii.ox.ac.uk

Tulio de Souza
Oxford University Computing Laboratory
tulio.de.souza@comlab.ox.ac.uk

Ian Brown
Oxford Internet Institute
ian.brown@oii.ox.ac.uk

Abstract

We examine the problem of mapping internet filtering, or censorship, at a finer-grained level than the national, in the belief that users in different areas of a country, or users accessing the internet through different providers or services, may experience differences in the filtering applied to their internet connectivity.

In investigating this possibility, we briefly consider services that may be used by researchers to experience a remote computer's view of the internet. More importantly, we seek to stimulate discussion concerning the potentially serious legal and ethical concerns that are intrinsic to this form of research.

1 Introduction

Many nations around the globe participate in some form of internet filtering[3]. Whilst filtering and censorship can, to an extent, be open and transparent, their nature tends towards secrecy. In order to understand the extent and nature of filtering around the world, we desire the ability to experience directly the limitations imposed on these internet connections.

National-level filtering, however, is simply the crudest form of such mapping. Whilst many states have national filtering policies, there is some evidence that the specific implementation of these may vary from region to region, from ISP to ISP and even from computer to computer. In order to fully understand filtering and its role in the globally networked world, it is extremely useful to explore connectivity at a more geographically and organisationally fine-grained level.

To this end, it is desirable to experience the Internet as viewed by a computer in a location of interest. There are a number of existing services specifically designed to allow this: VPN software and proxy services are well-known tools to allow a remote computer to route through a given remote network, and the well-known Tor anonymising network provides a similar service specifically aimed at bypassing national-level filtering.

For the purposes of wide-scale research, however, many of these services are relatively rare and require explicit access. Further, many of these services are employed directly to avoid filtering and thus to allow filtered users to access unfiltered connections. Clearly, such a service is less likely to exist on heavily filtered connections. In deliberately investigating filtered connections, it may be necessary also to explore other forms of information.

2 Motivation

There are many technical approaches to internet filtering employed around the world, applied to a greater or lesser extent. The most well-known filter is almost certainly China's "Golden Shield" (金盾工程, *jīndùn gōngchéng*), commonly known as the "Great Firewall of China", which represents arguably the largest and most technologically advanced filtering mechanism in use today.

Despite the technological sophistication of the Chinese national firewall, it is subject to a number of limitations. With a population of roughly 1.3 billion and an internet penetration rate estimated at almost 32%, the number of Chinese internet users is comparable to the combined populations of the US and Mexico. At such a scale economies must be made in the mechanisms of filtering to reduce the required resources to a manageable level. An excellent study of the technology underlying the Chinese national firewall was presented by Clayton et al[2].

Many other countries, however, perform internet filtering with significantly lower budgets and technical investment. Technologies range from crude blocking of large portions of the internet, to sophisticated and subtle blocking of specific content. A global view of internet filtering has been comprehensively presented in [3]. This work is

notable not just for its scope, but for its focus on the sociological as well as technical aspects of filtering, covering the nature of filtered topics and the levels of state transparency in the filtering process.

At a national level, however, filtering beyond crude mechanisms is often considered infeasible due not only to computational, but also to the organisational requirements of such systems; even if sufficient technological resources are available, the dynamic nature of the internet imposes a significant administrative burden in maintaining up-to-date filtering rules.

In solving this second problem states may choose to provide broader filtering guidelines to be implemented by local authorities or individual service providers, resulting in potential differences between the filtering experienced between users in different geographical locations or those using different providers. It is also possible, and has been observed in a number of cases, that a state may deliberately choose to restrict internet services to a greater or lesser extent in certain locations as a result of unrest or disaster.

To understand the technologies employed by states in filtering the internet, and the decisions behind this filtering, we therefore see great interest in studying the extent and nature of filtering at a regional and organisational, rather than national, level. We believe that this will provide a much more sophisticated picture of filtering around the globe, and provide a valuable source of information for internet researchers.

3 Filtering Technologies

The development of the internet was neither carefully planned, nor accurately predicted. It has expanded through the accretion of protocols, services and applications that have been extended and improved far beyond their original purpose. As such, many of the protocols provide opportunities both for filtering technologies, and for attempts to bypass or study those technologies.

There are a number of methods applied to filter internet connections at a national level. These have been usefully categorised by Murdoch and Anderson[7] as follows:

- **TCP/IP Header Filtering:** IP, the Internet Protocol, is the fundamental protocol by which traffic passes across the internet, encoded in IP *packets*. Filtering may occur via inspection of the *header* of an IP packet, which details the numerical address of the packet's destination. Packets may therefore be filtered according to lists of banned destination IP addresses. This method is simple and effective, but difficult to maintain due to the potential for services to change, or to have multiple, IP addresses. This approach may also incur significant "collateral damage" in the case of services that share IP addresses, causing multiple innocent services to be blocked along with the desired target.
- **TCP/IP Content Filtering:** Rather than inspecting the header, a filter may search the content of traffic for banned terms. This is a far more flexible approach to filtering, allowing packets to be blocked only if the include banned keywords or the traffic patterns of particular applications. This approach is also known as *deep packet inspection*, and is known to be employed to some extent by the Chinese national firewall. Deep packet inspection can be partially defeated by using encrypted connections, however filters may choose simply to block all encrypted connections in response, or to block traffic according to identifying traffic signatures that can occur even in encrypted protocols. The most significant limitation of this approach is that inspection of traffic content comes at a significant computational cost.
- **DNS Tampering:** The DNS protocol maps human-readable names to IP addresses on the internet, and is thus critical for most user-focused services such as the web. By altering DNS responses, returning either empty or false results, a filter can simply and cheaply block or redirect requests. This mechanism is simple to employ and maintain, but limits filters to entire websites, and can be relatively easy to bypass for technical users. This approach is employed by, among others, the Turkish state when blocking websites.
- **HTTP Proxy Filtering:** A more sophisticated approach is to pass all internet traffic through an intermediary "proxy" service that fetches and, typically, caches information for users. This is a common internet service that can be used to speed up internet connections and reduce traffic. A suitably enabled proxy can, however, employ sophisticated filtering on certain destinations, whilst leaving other connections alone. This approach can, by ignoring the majority of traffic, be efficient on a national scale while still allowing for detailed filtering similar to TCP/IP content filtering.
- **Other Approaches:** A variety of other means can be taken to regulate content on the internet. States can request that websites are removed from the internet, either by taking down their servers or by removing their names from the global DNS records. A state may also choose not to block a connection entirely, but to slow any connection to that site to unusable levels. At a less technical level, legal and

social constraints can be imposed to may accessing certain services illegal or socially unacceptable.

It has been noted, in [3] that many states begin by employing IP header filtering before moving on to more sophisticated methods as citizens protest the limiting of their connections. In the case of sophisticated national-level connections it is likely that a combination of these methods will be employed in order to meet the various constraints of large-scale filtering.

4 Mapping Filtering

A number of projects exist that provide insight into internet censorship around the world, both from the perspective of learning which sites are filtered and from the more practical approach of bypassing filtering. The most thorough study of global internet filtering is from Deibert et al[3], who present an in-depth global study of tools and techniques of filtering. The related Herdict project[11] allows users to report apparently blocked websites, via a browser plugin, to build up a global map of filtered sites. The Alkasir project[1] combines user-based reporting of blocked content with an anti-censorship tool that attempts to penetrate such filtering.

In bypassing internet filtering, the most well-known technology is the Tor project[4], which allows users to reroute their connections through a global network of volunteer-run anonymising proxy servers. This network, originally designed to preserve the connection-level privacy of users, was found to be an excellent tool for bypassing national filtering and now invests significant resources in supporting this use. Similar tools include Psiphon[8] as well as numerous Virtual Private Network (VPN) servers that allow users to evade national filters. All of these services work in a similar manner: by rerouting a connection through a server located in a different country, the user experiences the internet as if their connection originated in that country. Thus, a user from Saudi Arabia can route their connection through a US computer and bypass all filters run by their state, at the cost of some slowing of their connection and gaining those filters, if any, imposed by the US.

From these examples, we can observe two major possibilities for studying internet filtering. The first is to ask users in a given country to report their experience, as exemplified by the Herdict project; the second is to make use of an available service, such as a Tor node, in that country to experience the filtering directly. Both of these approaches have limitations that we explore in detail below.

Fundamentally, both of the aforementioned approaches suffer from a lack of availability that we see no easy way to avoid. In requesting users to directly report their experiences, Herdict relies on reaching interested and informed users. Tor relies on technically knowledgeable users to set up relays that require both significant resources and a willingness to face potentially serious legal issues[10]. In particular, at time of writing the Tor network does not report any publicly available servers in China[1].

The advantage of using a system such as Tor, Psiphon or VPN services is that they allow a researcher directly to control the flow of traffic. Sites of interest and even specific patterns of traffic can be directly sent and examined. This allows for a much more detailed examination of the technical measures employed on a given network. The approach taken by Herdict, however, cannot currently reproduce this level of sophistication. In the absence of a large network of experienced and technically capable users. user-level reporting only provides that a site appears to be unavailable, without reference to the conditions that cause the unavailability[2].

In order to achieve the fine-grained mapping of filtering that we desire, there are two major points of interest beyond those commonly considered by the most well-known current mapping projects. The first of these is the precise geographical location of a particular computer. The ability to determine the originating country of an IP address is relatively well known, and location to the level of an individual city can be achieved with some accuracy. Recent results[12] have proposed mechanisms that achieve a median accuracy of 690 metres, albeit within the US. This simple extension, we propose, would provide a valuable source of data on the applications of filtering. In many cases it is also possible to determine which organisation has been allocated any particular IP address, to the level of an ISP or major company. Both of these pieces of information can be used to build up a much more detailed view of filtering.

The second point of interest is to study, in detail, the technical nature of the filtering that is imposed on a given connection in a given location. While work has been conducted into specific methods, as in the work of Clayton et al. relating to the Chinese national filter, most large-scale projects appear to be focused more on the existence of filtering rather than the details of its implementation.

4.1 Extending Reporting Approaches

The approach taken by the Herdict project, which relies on volunteer participation to gather data, can be highly

[1] Specifically, there are no announced *exit nodes*, which would be the most feasible way to examine network filtering, reported as located on the Chinese mainland.

[2] The Herdict project does allow a user to express their opinion as to the cause of the blocking, but in the absence of direct experimentation this data has significant limitations.

effective if sufficient volunteers can be found. Herdict currently provides a webpage that attempts to direct a user's browser to load a random potentially-blocked site, and to report their experience. The project also makes available a web browser plugin that allows users to report sites that appear blocked. By focusing on the web browser environment, Herdict greatly reduce the effort required for user participation. The importance of this approach to usability, and the trust implicitly gained through the familiarity of the web browser, should not be overlooked.

This volunteer approach could naturally be extended to the use of more sophisticated tools to detect the presence of filtering automatically and, where possible, test the mechanisms employed. The detection of DNS filtering, IP blocking and even deep packet inspection is often simple enough in itself, particularly when the results of requests can be compared against reference requests made in other countries. It is, however, much more difficult to discover specifics of filtering mechanisms without direct, interactive access to the filtered network connection.

Our own experiments have resulted in a simple application that can detect a number of basic types of filtering, and has been tested on our own servers against deliberately filtered IP ranges and poisoned DNS responses. We make use of the freely-available MaxMind GeoIP database[6] to resolve IP addresses to the city level with a tolerable level of accuracy. At this point, however, our research has been limited, in part due to ethical concerns that we detail below, to proof of concept experiments for which we do not have useful results to present.

A dedicated application to detect and categorise filtering allows for a much higher level of accuracy with respect to the nature of reported filtering. Whether an effective number of users could be persuaded to run such an application is another matter. Therefore, while a standalone tool to map filtering would offer great flexibility, the barrier to entry for volunteers is potentially too high. Browser-based environments, such as JavaScript or Java applets, are likely to strike a useful balance between power and ease for end-users.

It is worth noting the Switzerland tool[9] developed by the Electronic Frontier Foundation, that aims to detect ISP-level filtering of peer-to-peer applications and violations of network neutrality principles. This tool detects many forms of network manipulation applied to an end user's connection, and offers the potential to be adapted for the purposes discussed here.

4.2 Direct Information Sources

As we have seen above, obtaining direct access to filtered connections is desirable for maximum flexibility. This can be achieved through Tor, Psiphon or open VPN services, all of which are specifically design to route traffic for third parties. Although some restrictions may exist on the access to these services, they provide an excellent platform for examining filtering when available. We note above that China does not appear to have any available Tor nodes; many other nations that reportedly engage in significant filtering, which are thus of greatest interest, show similarly low availability of such services. Despite the size and success of the Tor network in achieving its goals of anonymity and anti-censorship, this lack of availability limits its use for mapping global filtering. Where available, however, it is arguably the most powerful tool available to us. Similar services to Tor, including open VPNs, suffer from similar lack of scale to a far greater extent.

It is worth considering, therefore, if common services exist that allow for indirect exploration of filtering. The most obvious of these are DNS servers; these are widely available across the internet, often as an open service available to any users that choose to connect to them, and run a distinctive service that can be easily discovered. Their involvement in one of the major types of filtering, namely DNS poisoning, makes this particular type of filtering trivial to detect across much of the globe – one can simply connect to a DNS server in a locality where filtering is suspected and make DNS requests. If inconsistent results are found then these can be compared against reference requests from a trusted, non-filtered DNS server.

There are a small number of other well-known internet services that can be made to relay connections for a third party, although these are not typically common enough to allow for broad-scale research. Certain IRC servers, open shell access through telnet or SSH, open mail relays and various others offer the potential, however their scarcity and the difficulty of discovery make them a poor avenue of enquiry.

If we consider more legally and ethically questionable methods, there are a number of protocols that have the potential to be "repurposed" for the detection of filtering. Peer-to-peer filesharing networks result in large networks of home PCs running services that are accessible from any computer and that are themselves designed to connect to, and relay for, third parties. While these are unlikely to offer the flexibility of services such as Tor, there are several protocols, such as BitTorrent, that are amenable to this form of information gathering. It is worth highlighting at this point that such deliberate misuse of a service is likely to fall foul of the law in many jurisdictions, whilst simultaneously opening the operator of the service to potential repercussions if their connection is detected attempting to access banned content.

We find it impossible to resist mentioning a possibility open to those willing to throw law and ethics aside

entirely: many modern computer viruses exist solely to create networks of infected, or "zombie", PCs that can be entirely controlled from a central location. These captive systems are typically used, for the benefit of organised criminals, to send high volumes of spam emails or to blackmail organisations through denial-of-service attacks on their networks. Gaining access to such a botnet, some of which have been known to comprise tens of million PCs distributed across the globe[5], would provide an impossibly rich platform for these, and many other, network experiments.

5 Ethics and Legality

While many technical approaches, and challenges, exist for mapping global filtering, there are a number of serious legal and ethical issues to be faced with performing this research.

We have already mentioned that deliberate misuse of a network service may be illegal in many jurisdictions, and such misuse without a user's consent may well be considered unethical. Even when using openly available and general-purposed services, however, there are serious considerations when attempting to access blocked content via a third party.

In many situations, a user is unlikely to face repercussions for being seen to be attempting to access blocked content. The scale of internet use, even in smaller countries with low internet penetration rates, is simply too high for there to be serious policing of users who request filtered content. It is likely that, in the vast majority of cases, such attempts may not be logged at all. However, users in specific contexts may be put at risk.

The legality of attempting to access filtered content is also a concern. Many nations have somewhat loosely-defined computer crime laws, and often prefer to prosecute crimes involving computers under existing legislation rather than through creation of new laws. The legal status of attempting to access blocked content, however, and of attempting to bypass such blocks is not something a researcher can afford to ignore.

From the point of view of a researcher, these concerns are exacerbated by two factors: the concentrated attempts to access filtered content that is caused by a detection tool, and the wide variety of laws and social conventions that exist around the globe.

By their nature, the filtering detection mechanisms that we have discussed, and any that we can feasibly imagine, detect filtering by attempting to access filtered content: by requesting websites or IP addresses that are known, or are believed or likely, to be banned. As we have stated above, it would be largely impractical for a state to take note of every blocking action taken by their filter. It is possible, however, that sufficiently high-volume requests for banned content may be considered worthy of further action. A user innocently aiding a researcher in mapping their national filter, resulting in their computer suddenly attempting to connect to all forms of banned content, may find themselves under very unwelcome scrutiny.

It is also of great concern that a researcher not cause a user to unwittingly break the law with respect to the content that they direct a user to access. With the wide global variance in law, great care would have to be taken that a censorship tool not attempt to access content that was directly illegal. Pornography, particularly with respect to those under the local age of legal consent, *lèse majesté* and insults to religion are all sensitive issues that vary widely between cultures.

Volunteers that participate in research of this nature by running a filtering detection tool must do so having been fully informed as to the nature of the tool and the potential risks involved. From this perspective there is a significant added burden on the researcher to state to the participant, who may well not have any significant level of technical expertise, what the tool will do and what particular risks they run.

In the case of relay services, such as Tor or Psiphon, consideration must be given to the safety and security of the user operating the service. Due to their nature these services are frequently abused, and operators of such services must be prepared to defend their operation of the service. The Tor Project, in particular, invests significant efforts in education both for operators and for users. This does not, however, reduce the burden on a researcher taking advantage of such a service to ensure that they do not harm or endanger the operator through their actions.

6 Conclusions

We propose that it is in general false to consider internet filtering as an homogeneous phenomenon across a country, and that the practicalities of implementing a filtering regime are likely to result in geographical and organisational differentiation between the filtering experienced by users. We believe that the study of these differences are of great interest in understanding both the technologies and the motivations behind filtering, and propose a number of mechanisms that could be employed to gain this understanding.

However despite the existence of a number of technological and social avenues to aid in this research, we see a number of serious legal and ethical concerns that must be thoroughly considered in order to undertake broad-scale research of this nature. Beyond the more obvious pitfalls of misusing third-party services in an attempt to conduct this research, there are more subtle issues. The necessity

of attempting to access blocked content, and the legality and ethics of performing this via a third-party volunteer or service operator are all worthy of serious discussion by researchers in this field.

Despite these concerns, and the technical hurdles to gaining a detailed picture of global internet filtering, we consider that research into this subject presents a number of interesting problems, and can provide insight into the development of the internet and its ongoing social and political role both the national and international level.

References

[1] W. Al-Saqaf. Alkasir for Mapping and Circumventing Cyber-Censorship. `http://www.alkasir.com/`. Accessed May 8th, 2011.

[2] R. Clayton, S. J. Murdoch, and R. N. M. Watson. Ignoring the great firewall of china. In *In 6th Workshop on Privacy Enhancing Technologies*. Springer, 2006.

[3] R. J. Deibert, J. G. Palfrey, R. Rohozinski, and J. Zittrain. *Access Denied: The Practice and Policy of Global Internet Filtering (Information Revolution and Global Politics)*. MIT Press, 2008.

[4] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The Second-Generation Onion Router. In *Proceedings of the 13th USENIX Security Symposium*, August 2004.

[5] Matt Thompson. Mariposa Botnet Analysis. Technical report, Defence Intelligence, 2009.

[6] MaxMind Inc. MaxMind GeoIP City Database. `http://www.maxmind.com/app/city`. Accessed May 8th, 2011.

[7] S. Murdoch and R. Anderson. Tools and Technology of Internet Filtering. In R. Deibert, editor, *Access Denied: The Practice and Policy of Global Internet Filtering (Information Revolution and Global Politics Series)*, chapter 3, pages 57–72. MIT Press, 2 edition, Dec. 2008.

[8] Psiphon Inc. The Psiphon Project. `http://www.psiphon.ca/`. Accessed May 8th, 2011.

[9] The Electronic Frontier Foundation. Switzerland Network Testing Tool. `https://www.eff.org/testyourisp/switzerland`. Accessed May 8th, 2011.

[10] The Electronic Frontier Foundation. Tor Project Legal FAQ. `https://torproject.org/eff/tor-legal-faq.html.en`. Accessed May 8th, 2011.

[11] The Herdict Project. The Herdict Project. `http://www.herdict.org/`. Accessed May 8th, 2011.

[12] Y. Wang, D. Burgener, M. Flores, A. Kuzmanovic, and C. Huang. Towards Street-Level Client-Independent IP Geolocation. In *NSDI*. USENIX Association, 2011.